

Visualization of uncertainty associated with spatial prediction of continuous variables using HSI color model: a case study of prediction of pH for topsoil in peri-urban Beijing, China

TAN Man-zhi^{1,2}, CHEN Jie^{1*}

¹ State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, P.R. China;

² Graduate School of the Chinese Academy of Sciences, Beijing 100049, P.R. China

Abstract: Hue-Saturation-Intensity (HSI) color model, a psychologically appealing color model, was employed to visualize uncertainty represented by relative prediction error based on the case of spatial prediction of pH of topsoil in the peri-urban Beijing. A two-dimensional legend was designed to accompany the visualization—vertical axis (hues) for visualizing the predicted values and horizontal axis (whiteness) for visualizing the prediction error. Moreover, different ways of visualizing uncertainty were briefly reviewed in this paper. This case study indicated that visualization of both predictions and prediction uncertainty offered a possibility to enhance visual exploration of the data uncertainty and to compare different prediction methods or predictions of totally different variables. The whitish region of the visualization map can be simply interpreted as unsatisfactory prediction results, where may need additional samples or more suitable prediction models for a better prediction results.

Keywords: Hue-Saturation-Intensity; color model; spatial prediction; uncertainty; visualization

Introduction

Since 1980s, mathematical model has been broadly applied in spatial predictions of soil and other landscape attributes. Soil is a very complicated non-linear dynamics system on the earth surface. Any model can not absolutely, accurately describe and reflect soil variability in the real world. Therefore uncertainty is always being in spatial prediction results. Spatial uncertainty has two connotations, namely attribute uncertainty and spatial location uncertainty, existing in all links, for example, datum measurement, statistical calculation and construction of model. Furthermore, uncertainty of original datum and model can be propagated through spatial predicted model, inducing uncertainty

of final output results and the related error appraisal (Mowrer and Congalton 2000). Since 1990s, it has been attached more importance to assessment of the uncertainty in spatial prediction of soil properties, and this assessment has being brought into subsequent decision-making processes, such as delineation of polluted areas or identification of zones that are suitable for crop growth (Goovaerts 2001). The logical standard judging quality of soil spatial predicted map is the content of information that consists of information content of soil properties, uncertainty analysis, and information recognition potential. Among the above information, the visualization of uncertainty is one of key factors to determine quality of soil spatial predicted map. With the developing of GIS technology, visualization of uncertainty is becoming a vital new connotation of soil mapping process. Visualization of uncertainty can be divided into two different ways: 1) independent visualization of uncertainty, as an independent adjacent map for spatial prediction map; 2) synchronous visualization, combined with uncertainty analysis and spatial predicted results on the same map.

Until the mid-1990s, uncertainty assessment has been essentially performed using non-linear kriging approaches that aim at evaluating the probability for the target attribute greater than the threshold value and setting confidence range at a specific unmonitored location (Smith et al. 1993; Webster and Oliver 1989; Goovaerts and Journel 1995; Li et al. 2004). Synchronous visualization is always implemented through two ways. One new layer with the uncertainty information is established on the spatial prediction map, which changes information of uncertainty mainly through transparency, haze or fog, blurring and texture and patterns under spatial prediction map (Dutton 1992;

Project Foundation: Under the auspices of Knowledge Innovation Frontier Project of Institute of Soil Science, Chinese Academy of Sciences (No. ISSASIP0716) and the National Nature Science Foundation of China (No. 40701070,40571065)

Received: 2008-02-22; Accepted: 2008-03-28

© Northeast Forestry University and Springer-Verlag 2008

The online version is available at <http://www.springerlink.com>

Biography: TAN Man-zhi (1978-), female, a native of Wangjiang of Anhui Province, Ph.D. and Assistant Professor, specialized in spatial variability and uncertainty analysis. E-mail: mzhtan@issas.ac.cn

*Corresponding author: Chen Jie (1966-), Ph.D and Professor.

E-mail: jchen@issas.ac.cn

Responsible editor: Hu Yanbo

Maceachren 1992; Monmonier 1990; Pang et al. 1994). The other constructs spatial prediction map with uncertainty information through improvement of mapping technology combining the uncertainty information with predicted results in the same layer. Hengl et al. (2004) used HSI color model to visualize both the spatial prediction of topsoil thickness and uncertainty analysis in the same layer for the first time. However, the related researches in the above field are scarce in China. This paper uses HSI color model to visualize spatial predicted result associated with uncertainty expressed by relative prediction error on the same map, based on a case study of spatial prediction of topsoil pH in Beijing Peri-urban.

Materials and methods

Sample collection and analysis

The study area is located within the peri-urban zone of Beijing (about 2 600 km²). In total of 220 topsoil samples (0–20 cm) were collected from the grid cells about 2 km × 2 km from April to May of 2001. Each point in Fig. 1 (enlarged diagram on the right side) represents a 5 m × 5 m block where five soil samples were collected from the center and corners to make one composite sample. The global positioning system (GPS) method was used to record geographical coordinates of each sampled site, and environmental observations were described during the fieldwork. The values of pH were measured by electrode method in liquor mixed water with soil by 2.5:1.

Geo-statistical method

The semi-variance function of sampling point was calculated based on GS + software. The pH values of un-sampled soil were best and un-partially predicted through ordinary Kriging method in Geo-statistical method by ArcGIS software, and prediction map and standard predicted error map were derived through the software.



Fig. 1 Distribution of soil-sampling sites in Study area

The HSI color model

The first important requirement for successful visualization of uncertainty is to select a psychologically appealing color model. For example, a logical color variable for visualizing uncertainty is whiteness. HSI model seems to be the most promising for the

visualization of uncertainty. This was confirmed by Jiang (1996) who conducted a number of perception tests. Another important reason for using this color model is the fact that it is geometrically related to the RGB color cube, which makes it possible to calculate intermediate and mixed colors. Whiteness is amount of the white color, which in the color cube can be visualized as the shortest distance to the white corner of the RGB cube. The hue (H) represents the visual sensation of the color type, and is calculated as the number of degrees around the axis. The saturation (S) represents the degree to which the color expresses its hue, and is calculated as the radial distance from the diagonal axis. The intensity (I) represents the visual sensation of brightness (Fig. 2a).

Visualization of uncertainty

The uncertainty of prediction is represented with the normalized or relative prediction error expressed in percentage, which is to divide the prediction error of the transformed variable by the standard deviation of observed samples.

$$\sigma_{E,r}(s_0) = \frac{\sigma_E(s_0)}{S_z} \times 100\% \quad (1)$$

where, $\sigma_E(s_0)$ is the prediction error, S_z is the standard deviation of observed samples. A satisfactory prediction is when the model explains more than 85% of the total variation. As a rule of thumb (Jiang 1996; Hengl 2003; Park and Vlek 2002), we can consider that a value of $\sigma_E(s_0)$ close to 40% means a fairly satisfactory accuracy of prediction. Otherwise, if the values get above 80%, the model accounted for less than 50% of variability at the validation points and prediction is unsatisfactory.

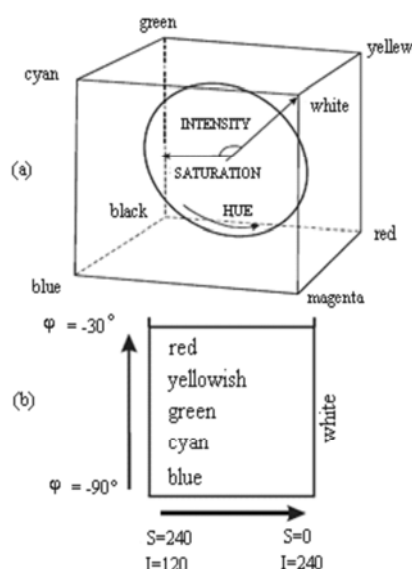


Fig. 2 Design of the legends for visualization of uncertainty. (a) Hue-Saturation-Intensity color model and the Red-Green-Blue color cube; (b) the two-dimensional legend used for visualization of uncertainty in quantitative data

The values of prediction and relative prediction error must be standardized by maximal and minimal value before coding by using:

$$z_r = \frac{\hat{z} - z_1}{z_2 - z_1}; u_r = \frac{\sigma_{E,r} - u_1}{u_2 - u_1} \quad (2)$$

where, z_r is the standard prediction, u_r is the standard prediction uncertainty ($z_r, u_r \in [0, 1]$). \hat{z} is the prediction map, $\sigma_{E,r}$ is the relative prediction error map standardized using Eq.(1), z_1 and z_2 , u_1 and u_2 are the lower and upper limits of the predicted values and relative prediction error. For prediction variables, the range limits are the measured minimum and maximum, and for uncertainty the thresholds are 40% and 100%.

The map of predictions and uncertainty can be visualized simultaneously by coding the predictions with hues and uncertainty with whiteness. The predictions are coded to hue using:

$$\phi_1 = -90 + z_r \times 300 \quad (3)$$

$$\phi_2 = \begin{cases} \phi_1 + 360 & \text{if } \phi_1 \leq -360 \\ \phi_1 & \text{if } \phi_1 > -360 \end{cases} \quad (4)$$

where, ϕ_1 is the hue angle in degrees measured clockwise, ϕ_2 is the value transformed from the -360 to 360 range. The HSI-coded image is then derived using:

$$H = (\phi_2 + 360) \cdot \frac{240}{360} \quad (5)$$

$$\begin{aligned} S &= (1 - u_r) \cdot 240 \\ I &= (1 + u_r) \cdot 120 \end{aligned} \quad (6)$$

Design of the legend

Two-dimensional legend was designed for prediction of continuous variables. Here, vertical axis represented the predicted values and horizontal axis (whiteness) is used to visualize the prediction error (Fig. 2b). On the vertical axis, hues ranged from blue for low values (-90°) to red for high values (-30°). A part of the hue circle representing magenta (-30° to -90°) has been omitted to avoid confusion between high and low values. On the horizontal axis, the saturation-intensity changes linearly from low to total whiteness. The combined visualization gives insight into the relationship between uncertainty and input data for the given thresholds. Full color (the original RGB) means the minimum uncertainty is equal to or less than 40%, and full whiteness indicates the maximum uncertainty equal 100%.

Results and discussion

Descriptive statistics

Descriptive statistics results showed that the variability of pH was quite small. Distribution of the data was accordance with normal distribution using K-S test (K-S, 0.362). The mean value was 8.37 (the minimum was 7.35; the max. 9.06), therefore the soil was alkaline.

Geo-statistical analysis

Experimental Isotropic semi-variogram of topsoil pH at full extent was matched with exponential model best, R^2 was 0.935. Effective range was 18.99 km, Nugget was 50.2%, demonstrating that pH had a moderate spatial correlation in a relative large range.

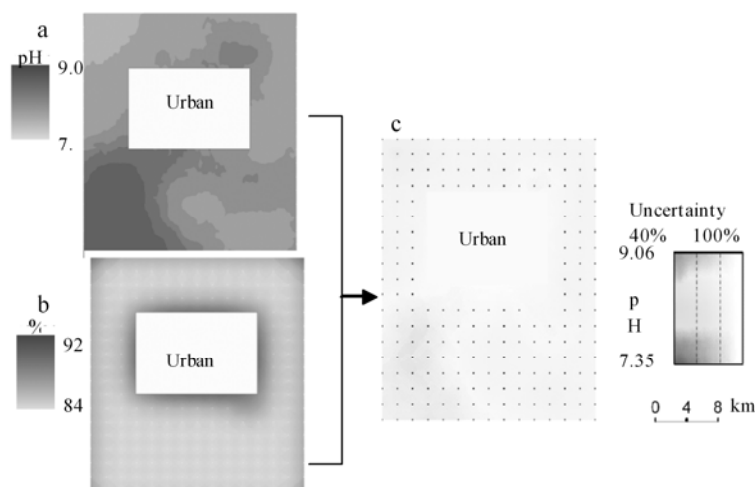


Fig. 3 Prediction map of topsoil pH (a), relative prediction error (uncertainty) (b), visualization of both predictions and prediction uncertainty of pH for topsoil, which accompanying two-dimensional legend—predictions with hue, uncertainty with whiteness (c)

Visualization of predictions and uncertainty

Spatial prediction map of pH was derived by ordinary Kriging method in ArcGIS software. The results of cross validation showed that mean standardized error was -0.0022, root-mean-square error (0.1811) nearly equaled to average standard error (0.1736). The value of root-mean-square standardized error approximately equaled to 1. The mean error was -0.0005. All data demonstrated that the precision of prediction was quite perfect.

The spatial prediction map of pH presented patch shape. High values were mainly distributed in the west-south corner of the study area. Outside of the west-south corner, the value was decreasing. Low values were distributed in the west-north corner (Fig. 3a). Uncertainty map showed that relative prediction error was increasing far from the sample points. In the center and four corners, the uncertainty was relatively higher, of which the highest value was 92% (Fig. 3b). Most area was whitish (Fig. 3c), meant that the spatial prediction was not satisfactory. The visualization resulted in most of the study area far from the points being pale. These areas would probably need additional samples or more suitable prediction models. This study used different kriging methods to do the uncertainty analysis and the results were nearly the same, the uncertainty was all very high. In case of significance was very little, this paper did not list other results. If the interpolated data would be perfect, it is very interesting to do this.

Conclusions

The results of cross validation showed that precision of spatial prediction was relatively high. But the uncertainty was also quite high, which the mean relative prediction error was about 88%. Prediction standard error was influenced by sampling density and distribution pattern. The relative prediction error, i.e., the uncertainty will be smaller, and the spatial prediction will be more credible, if the sampling density is denser and the distribution is more regular. In this case the samples distributed uniformly, but the sampling space was large, and relative prediction error was quite high. Therefore it is an effective way to improve the sampling density to reduce the prediction uncertainty.

Visualization of both predictions and prediction uncertainty offers a possibility to enhance visual exploration of the data uncertainty and to compare different prediction methods or predictions of total variables. The whitish area can be simply interpreted as unsatisfactory predictions, i.e. areas that need to be sampled additionally.

In the process of visualization both prediction and prediction uncertainty at the same time, the two dimensional legends showed non-neglected limiting effect. The matter is that it is not easy to match the colors on the HSI coded image with the legend because the pale colors are harder to be distinguished. Moreover,

it is not easy to read the actual uncertainty within $\pm 5\%$ of the relative prediction error by just looking at the whiteness on the map. Therefore, further effort should be focused on improvement of the legend designed.

Analysis and visualization of uncertainty don't aim at eliminating uncertainty in spatial prediction, but using existing information to assess the accuracy and precision of uncertainty models, and then visualize it straightly. It is very important to make actual decisions based on the prediction results. Take example of this case, the spatial prediction uncertainty is high, therefore it can't guide agriculture production and administer soil resource for fear acidification.

References

- Dutton G. 1992. Handling positional uncertainty in spatial databases. In Proceedings of 5th International Symposium on Spatial Data Handling. University of South Carolina, August 1992, pp460–469.
- Goovaerts P. 2001. Geo-statistical modeling of uncertainty in soil science. *Geoderma*, **103**: 3–26.
- Goovaerts P, Journel AG. 1995. Integrating soil map information in modeling the spatial variation of continuous soil properties. *Eur J Soil Sci*, **46**: 397–414.
- Hengl T, Heuvelink GBM, Stein A. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, **120**: 75–93.
- Hengl T. 2003. Visualisation of uncertainty using the HSI colour model: computations with colours. 7th International Conference on Geo-Computation (CD-ROM), University of Southampton, Southampton, p8.
- Jiang B. 1996. Fuzzy overlay analysis and visualization in geographic information systems, PhD Thesis, Utrecht: University of Utrecht.
- Li Yan, Shi Zhou, Wang Renchao, Huang Mingxiang. 2004. Estimates of electrical conductivity for coastal saline soil profile using cokriging under different sampling density. *Acta Pedologica Sinica*, **41**(3): 434–443.(in Chinese)
- Mowrer HT, Congalton RG. 2000. Quantifying spatial uncertainty in natural resources: theory and applications for GIS and remote sensing. Hannover: Ann Arbor Press, p350.
- Maceachren AM. 1992. Visualizing uncertain information. *Cartographic Perspective*, **13**: 10–19.
- Monmonier M. 1990. Strategies for the interactive exploration of geographic correlation. In: Proceedings of the 4th International Symposium on Spatial Data Handling. IGU, **1**: 512–521.,
- Park S, Vlek P. 2002. Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques. *Geoderma*, **109**(1-2): 117–140.
- Pang A, Furman J, Nuss W. 1994. Data quality issues in visualization. In: Robert J. et al. (eds), SPIE Vol. 2178 *Visual Data Exploration and Analysis*, p12–23. SPIE, February 1994.
- Smith JL, Halvorson JJ, Papendick RI. 1993. Using multiple-variable indicator kriging for evaluating soil quality. *Soil Sci Soc Am J*, **57**: 743–749.
- Webster R, Oliver MA. 1989. Optimal interpolation and isarithmic mapping of soil properties: VI. Disjunctive kriging and mapping the conditional probability. *J Soil Sci*, **40**: 497–512.